

The crystallization of asymmetric patchy protein models

Diana Fusco¹ and Patrick Charbonneau^{1,2,3}

¹*Program in Computational Biology and Bioinformatics, Duke University, Durham, NC 27708*

²*Department of Chemistry, Duke University, Durham, NC 27708*

³*Department of Physics, Duke University, Durham, NC 27708*

Asymmetric patchy particle models have recently been shown to describe the crystallization of small globular proteins with near quantitative accuracy. Here, we investigate how asymmetry in patch geometry and bond energy impact the phase diagram and nucleation dynamics of this family of models. We find the role of the geometry asymmetry to be weak, but the energy asymmetry to markedly interfere with the crystallization thermodynamics and kinetics. These results rationalize the success and occasional failure of George and Wilson’s proposal for protein crystallization conditions as well as provide physical guidance for developing more effective protein crystallization strategies.

PACS numbers: 87.15.ak, 82.70.Dd, 87.15.nt

Protein structures from X-ray and neutron crystallography are key to studying elementary biological mechanisms [1, 2]. The difficulty of obtaining diffraction-quality crystals thus severely limits our understanding of living systems. From a physical point of view, identifying successful crystallization conditions is equivalent to determining a protein solution phase diagram. Crystallizing a protein indeed typically involves placing a drop of protein solution near a high-salt aqueous buffer that drives the vapor diffusion of the solvent away from the drop. The non-volatile solutes then steadily concentrate and, if the initial conditions are properly chosen, a protein crystal forms [1]. The limited usefulness of existing physical descriptions and of knowledge-based approaches [3, 4], however, leave a vast space of solution conditions to be experimentally screened. A material understanding of protein assembly is thus essential to developing more effective crystallization strategies.

Soft matter descriptions of protein assembly based on particles with isotropic, short-range attractive interactions [5–7] – as suggested by early structural studies [8–10] – provide some conceptual guidance. They identify the region between the solubility line, above which the solution is stable, and the liquid-liquid critical point, below which the system gels [11], as the “crystallization gap” where crystal assembly is possible. This simplistic picture is, however, unable to reproduce many experimental trends [12–16]. The introduction of bond directionality in symmetric “patchy” models aims to better represent the effective protein-protein interactions that drive their crystallization [17, 18]. Yet the most commonly studied versions of these models have symmetric and interchangeable patches, which are atypical of real proteins [4, 19, 20] and insufficient to describe the assembly of even the simplest of globular proteins [21, 22]. In this letter, we thus investigate the role of patch geometry and bond energy asymmetry on the phase diagram and assembly dynamics of a family of protein models. Note that this additional anisotropy ‘direction’ also complements earlier experimental proposals for introducing richness to soft matter assembly [23] and opens the door to assembling

more complex structures in systems such as DNA-coated colloidal particles [24–26].

Model description – We consider patchy models hybrid between those of Sear [27] and Kern–Frenkel [28]. Hard spheres, whose diameter σ sets the unit of length, are decorated by three pairs i of patches that only interact with each other. Each pair’s properties are independently tuned [29]. The short radial extent of the square-well attraction, $\lambda_i = 1.1\sigma$ [30], and its surface coverage measured by the semi-opening angle of its conical segment, $\delta_i = \cos^{-1}(0.89)$, are chosen to be typical of protein-protein interactions [20, 22]. By contrast, the patch orientation and bond energy ϵ_i are randomly chosen under the constraint that the lattice formed by simply bonding the patches is that of the most commonly observed in monomeric protein crystals, the orthorhombic $P2_12_12_1$ [31, 32]. This lattice’s three non-intersecting two-fold screw axes guarantee a high number of rigid-body degrees of freedom with minimal symmetry constraints. We characterize the patches with energy and geometry asymmetry parameters

$$\zeta = \frac{(\epsilon_1 - \epsilon_2)^2 + (\epsilon_1 - \epsilon_3)^2 + (\epsilon_2 - \epsilon_3)^2}{2(\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2)}$$

$$\gamma = \frac{(I_1 - I_2)^2 + (I_1 - I_3)^2 + (I_2 - I_3)^2}{2(I_1^2 + I_2^2 + I_3^2)},$$

respectively, where I_i are the eigenvalues of the inertia tensor of the collection of equal-weight patches [29, 33]. Both ζ and $\gamma \in [0, 1]$, where 0 corresponds to an equal energy distribution ($\epsilon_1 = \epsilon_2 = \epsilon_3$) and cubically distributed patches, and 1 corresponds to a complete energy asymmetry ($\epsilon_1 = \epsilon_2 = 0$ and $\epsilon_3 = \epsilon_{\text{tot}}$) and a unit cell elongated in a single direction. The crystal symmetry group and the choice of λ_i , however, constrain γ to within $0 < \gamma \lesssim 0.1$ and prevent dimerization [29].

Phase diagram – For 30 randomly selected sets of patch geometry and bond energies we numerically determine the solubility line using free-energy integration and the metastable vapor-liquid line using Gibbs Ensemble Monte Carlo simulations [29]. The gas-liquid criti-

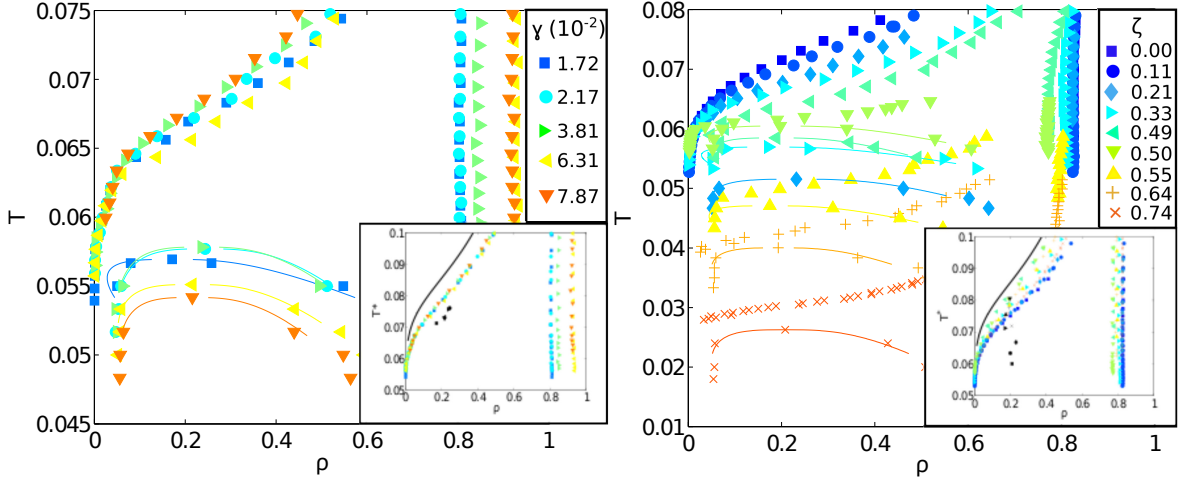


FIG. 1. Temperature-density ρ phase diagrams of patchy models with different γ at fixed $\zeta = 0.33$ (left panel), and with different ζ at fixed $\gamma = 0.0172$ (right panel). Insets depicts the same phase diagrams with T rescaled following WPT. The crystal-fluid coexistence lines (symbols) are then close to yet distinct from the WPT solubility line (solid black line). For visual clarity, only the critical points (black symbols) from the gas-liquid metastable lines (symbols plus line) are reported in the insets.

cal temperature T_c generally decreases with increasing γ because patch proximity anti-correlates bond formation and decreases the liquid entropy, although the limited number of systems studied partially hides this feature (Fig. 1) [34]. The solubility line, by contrast, is clearly similar for different geometries at fixed ζ and monotonically shifts to lower temperatures with increasing energy asymmetry (Fig. 1). In qualitative agreement with these results, the Wertheim perturbation theory (WPT) for associated liquids predicts that, independently of patch positions, $\beta\epsilon_{\text{tot}} + \beta\mu_{\text{bond}}^{\text{fluid}}$ is an invariant of temperature along the solubility line [30, 35–37]. WPT overestimates the solubility temperature at all densities ρ , but nonetheless remarkably collapses the simulation results (Fig. 1 insets) [29].

The numerical validation of WPT’s T_c predictions – accurate to within 10-15% – allows us to estimate the size of the “crystallization gap” for a broad variety of models (Fig. 2) [38]. Interestingly, we find that for patch energy sets $\{\epsilon_i\}$ giving a same ζ , a lower second virial coefficient B_2 results in a larger crystallization gap (Fig. 2A). Contrary to George and Wilson’s (GW) crystallization slot proposal that $\log(-B_2^*) < 5$ identifies facile crystallization [3], the asymmetric models reveal that B_2 does not by itself sets the size of the crystallization gap. The proposal is reasonable at low ζ , but breaks down at high ζ , where it even includes systems for which the critical point is fully stable (red star in Fig. 2A and C). In such a system access to the crystal from a slowly concentrating, low-density solution would have to side-step the metastable gas-liquid coexistence regime, which typically prevents the formation of all but the smallest crystallites [39, 40]. High interaction asymmetry therefore

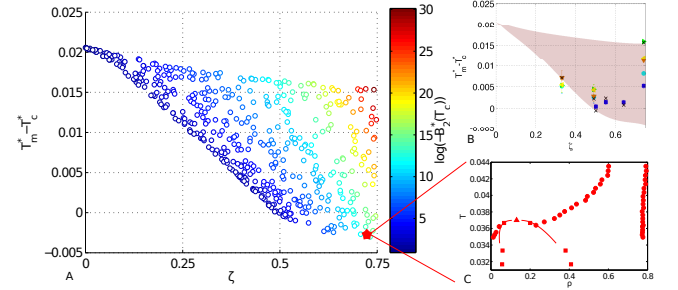


FIG. 2. A: WPT difference between rescaled melting temperature T_m^* and rescaled critical temperature T_c^* at the critical density $\rho_c \approx 0.2$, i.e. the “crystallization gap”. Each circle represents a distinct $\{\epsilon_i\}$ realization and is colored according to $B_2(T_c)$. B: WPT predictions (black crosses) compared with simulation results for different patch geometries (symbols as in Fig. 1). C: Phase diagram of a specific $\{\epsilon_i\}$ for which WPT predicts a stable gas-liquid coexistence. Even for an interaction range and patch coverage that would normally result into a metastable gas-liquid line, bond energy asymmetry can lift the critical point above the solubility line. An extrapolation of T_c using Grand Canonical Monte Carlo confirms this inversion [29].

provides a microscopic rationale for the failure of the GW proposal in a number of experimental systems [41, 42].

Crystallization – Even if crystallization is thermodynamically possible the free energy drive may be insufficient to induce a phase transition on experimentally-relevant timescales. We thus examine the role of asym-

metry on homogeneous nucleation using umbrella sampling simulations [29]. Systems near their critical density $\rho_c \approx 0.2$ at different degrees of supersaturation $\eta = \frac{T_m - T}{T_m - T_c}$, where T_m is the solubility temperature at that density, are considered. Unsurprisingly, the lower is the supersaturation, the higher is the free-energy barrier and the larger is the critical nucleus (Fig. 3B). Across various patch geometries qualitatively similar results are obtained, but increasing the energy asymmetry significantly lowers the chemical potential difference, $\beta\Delta\mu$, between the fluid and the crystal. At high bond energy asymmetry fewer patches dominate the energy of the two phases, which makes that contribution more similar and reduces the drive to crystallize. Higher densities are thus needed to obtain a comparable nucleation barrier. Although this effect is not a fundamental limitation for particles to crystallize, proteins in high-density solutions may partially unfold and aggregate, which interferes with crystallization [39]. In addition, at high ζ the narrow crystallization gap also results in larger free-energy nucleation barriers. High energy asymmetry thus hinders nucleation kinetics.

Classical nucleation theory (CNT) describes crystal reasonably well far above the critical point, but near and below T_c the assembly behavior is more complex. Previous studies of isotropically attractive systems have shown that well below T_c spinodal decomposition leads to dynamical arrest [11], because spontaneous density fluctuations result in dense regions within which binding is irreversible. In similar systems near the critical point, “two-step” nucleation is favored [7], because crystal formation is easier in high-density fluid regions than in low-density regions. The corresponding assembly behavior of patchy systems, whose low-density crystals may not be favored by spontaneous fluid density fluctuations [43], is here studied in unbiased constant NpT MC simulations. These simulations sketch out the minimum free energy path for the assembly, which we track along the largest drop and the largest crystal cluster reaction coordinates (Fig. 3) [29]. Far above T_c , the largest cluster is always crystalline and CNT applies. Near the critical point (within $\sim 10\%$ of T_c) a growing liquid drop first forms and only subsequent structural reorganization results in a large crystal cluster. Indeed, because of the low nucleation barrier within a dense fluid drop (Fig. 3B) the drop may shelter many microcrystals. A larger crystal then forms from the rearrangement of these small crystals, rather than from the growth of a single critical nucleus (red versus blue symbols in Fig. 3C), as is seen far above T_c . Yet, because these rearrangements are activated, forming a large cluster can be much slower than classical nucleation, which is rare but rapid (blue symbols in Fig. 3C). Below the critical point, the bonds are much less labile. The system forms a large dense drop, but no large-scale reorganization occurs on the simulation timescale. Only small crystal clusters are thus detected. These results reveal that patchy particles have an assembly dynamics comparable to that of isotropic

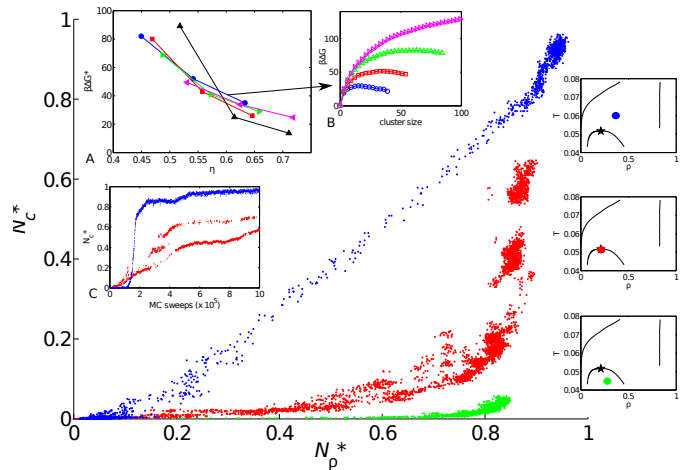


FIG. 3. Rescaled size of the largest liquid cluster N_ρ^* versus that of the largest crystal cluster N_c^* for a model with $\zeta = 0$ and $\gamma = 0.017$. The liquid cluster is rescaled over the number of particles in the system ($N=864$) and the crystal cluster over the size of the largest crystal cluster that fits in the simulation box (600). The different trends correspond to starting homogeneous fluid configurations under different conditions, as illustrated in the insets on the right. Above the critical point, nucleation barriers can be computed as shown in inset B (blue circle $\eta = 0.37$, red squares $\eta = 0.46$, green triangles $\eta = 0.55$ and magenta right-pointed triangles $\eta = 0.64$). Inset A illustrates the height of the nucleation barrier for different models (blue circles $\zeta = 0$ and $\gamma = 0.017$, red squares $\zeta = 0$ and $\gamma = 0.021$, green right-pointed triangles $\zeta = 0$ and $\gamma = 0.065$, black triangles $\zeta = 0.1$ and $\gamma = 0.017$, magenta left-pointed triangles $\zeta = 0.2$ and $\gamma = 0.017$.) Inset C provides the size of the largest crystal cluster along the simulation above (blue) and at the critical point (red).

particles, which confirms the relevance of the crystallization gap for identifying good crystallization conditions. If a large crystal is necessary for diffraction studies, one should target crystallization conditions well above the critical point, where CNT describes the dominant homogeneous nucleation pathway [43]. Recent advances in X-ray laser technology, however, provide diffraction data from relatively small crystals [44], which should extend the experimentally useful range of the crystallization gap closer to T_c , although probably still not below T_c .

Percolation – We finally consider direct percolation, which can also dynamically compete with crystallization. Below the percolation threshold $T_p(\rho)$ the system forms infinitely large spanning networks that can be long-lasting at high bond strengths [45]. To explore the interplay between percolation and bond energy asymmetry (patch geometry asymmetry has only a weak effect), we determine $T_p(\rho_c)$ using finite-size rescaling [29, 46] in systems with relatively low ($\zeta = 0.22$) and high ($\zeta = 0.55$) bond energy asymmetry (Fig. 4). In the first system the percolation threshold lies just above T_c , while in the sec-

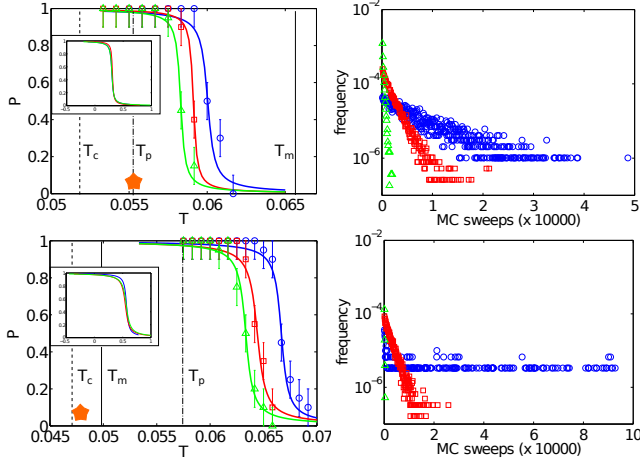


FIG. 4. Percolation behavior for $\zeta = 0.2$ (top) and $\zeta = 0.55$ (bottom). Left panels show the probability P of observing a spanning network as a function of temperature for system of size $N=2048$ (blue circles), 4000 (red squares) and 6912 (green triangles). The superimposed vertical lines indicate the critical temperature (dashed), the melting temperature (solid) and the percolation temperature (dot-dashed) estimated by finite-size scaling (inset). The right panels show the distribution of bond life time in the network respectively at $T=0.055$ and $T=0.048$ (orange stars in the left panels). Blue circles indicate the strongest bond, red squares the second strongest, and green triangles the weakest.

ond, in which the strongest bond is much longer-lasting, T_p is well above the solubility line. The dynamical relevance of percolation on crystallization is estimated from the distribution of bond lifetimes within the crystallization gap. At low bond energy asymmetry, the rearrangement of all bonds is observed within a few thousand MC steps. At high asymmetry, by contrast, the lifetime of the strongest bond (blue circles) is comparable to the length of the simulation (10^5 MC sweeps). The network is frozen and no rearrangement takes place. This observation suggests that identifying the crystallization gap may be insufficient for crystallizing particles with high energy asymmetry, because a long-lasting gel caused by direct percolation dynamically interferes with crystallization.

Conclusion – In order to gain insights into protein crystallization and soft matter assembly, we have considered the role of patch geometry and bond energy asymmetry on the crystal assembly of a family of schematic models. We find patch geometry asymmetry to have a weak effect, but bond energy asymmetry to severely impede the crystallization thermodynamics and kinetics. The crystallization gap shrinks, gel formation is favored, and nucleation shifts to higher supersaturations. The union of these observations suggests that to ease identifying the proper crystallization conditions, it is sometimes more effective to symmetrize the directional pair interactions between proteins, rather than specifically strengthen one of them, as is sometimes implicitly sug-

gested [4, 22]. It also offers a rationalization of the GW crystallization slot proposal as well as for its occasional failure. At low bond energy asymmetry, the B_2 slot prescription falls within the slot. At high asymmetry, a large crystallization gap is only observed for B_2 below the slot, which corresponds to long-living gels, while for B_2 within the slot the crystallization gap is very small or even negative. The GW crystallization slot is therefore a necessary but insufficient condition for detecting optimal experimental conditions.

Although we are now closer to understanding simple, monomeric protein crystallization, we are still a long way from capturing the assembly features of more complex proteins. Some proteins dimerize or change conformation on a timescale comparable to their crystallization, while membrane proteins typically require entirely different crystallization approaches than the one considered in this letter. Further modifications to patchy particle models, such as self-interacting or dynamically evolving patches, may thus guide our understanding of these complex and crucial biological objects.

ACKNOWLEDGMENTS

PC acknowledges NSF support No. NSF DMR-1055586.

Appendix A: Model details

The pair-wise interaction between particles 1 and 2, whose centers are at distance r_{12} apart, is

$$\phi(r_{12}, \Omega_1, \Omega_2) = \phi_{\text{HS}}(r_{12}) + \sum_{i=1}^n [\phi_{2i,2i-1}(r_{12}, \Omega_1, \Omega_2) + \phi_{2i-1,2i}(r_{12}, \Omega_1, \Omega_2)], \quad (\text{A1})$$

where Ω_1 and Ω_2 are the Euler angles. A hard-sphere (HS) potential captures the volume exclusion

$$\phi_{\text{HS}}(r) = \begin{cases} \infty & r \leq \sigma \\ 0 & r > \sigma, \end{cases} \quad (\text{A2})$$

where σ is the diameter of the particle. The patch-patch interaction is the product of a radial and an angular component

$$\phi_{2i,2i-1}(r_{12}, \Omega_1, \Omega_2) = \psi_i(r_{12}) \omega_{2i,2i-1}(\Omega_1, \Omega_2), \quad (\text{A3})$$

where

$$\psi_i(r) = \begin{cases} -\epsilon_i & r \leq \lambda_i \sigma \\ 0 & r > \lambda_i \sigma \end{cases}, \quad (\text{A4})$$

$$\omega_{2i,2i-1}(\Omega_1, \Omega_2) = \begin{cases} 1 & \theta_{1,2i} \leq \delta_{2i} \text{ and } \theta_{2,2i-1} \leq \delta_{2i-1} \\ 0 & \text{otherwise} \end{cases} \quad (\text{A5})$$

The range of the interaction λ_i is in units of σ , δ_{2i} is the semi-width of patch $2i$ and $\theta_{1,2i}$ is the angle between the vector r_{12} and the vector defining patch $2i$ on particle 1. An analogous definition holds for $\theta_{2,2i-1}$. For the purpose of this paper, λ_i and δ_{2i} are identical across patches and set respectively to 1.1σ , 0.47 rad.

Appendix B: Geometry asymmetry parameter

The geometry asymmetry parameter is defined as

$$\gamma = \frac{(I_1 - I_2)^2 + (I_1 - I_3)^2 + (I_2 - I_3)^2}{2(I_1^2 + I_2^2 + I_3^2)}, \quad (\text{B1})$$

where I_i represents the i^{th} eigenvalue of the inertia tensor of the object represented in Fig. B.1. Each patch (in red) carries a mass M at its center. The inertia tensor is computed over the collection of the weighted patches. The expression for γ guarantees the its value does not depend on the fictitious mass M and on the distance between the center of the patch and the center of the particles (radius of the particle) as long as these quantities do not vary from one patch to the other. Patches located on a perfect octahedron correspond to $I_1 = I_2 = I_3$, and consequently $\gamma = 0$.

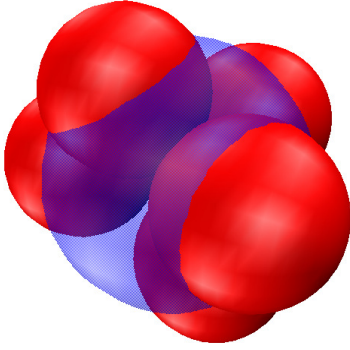


FIG. B.1. Sketch of the patchy particle. To determine the inertia tensor, we treat the particle as spherical balls (red) whose center is at the surface.

The patches are randomly distributed on the surface under the constraint that the simply bonded crystal belongs to the $P2_12_12_1$ symmetry group. As a consequence, patches cannot be too close to each other – otherwise bonded particles would overlap and the unit cell would stretch beyond the range of attraction $\lambda = 1.1\sigma$, which limits the achievable asymmetry and the maximum value

of $\gamma \approx 0.1$. Because a cubic symmetry ($\gamma = 0$ and limiting case of $P22_12_1$) is not realizable within the three screw axes symmetry of $P2_12_12_1$, γ is limited from below as well. The adopted range and width of the interaction and the $P2_12_12_1$ symmetry ensures that two particles can interact only through one bond. This condition, together with the impossibility for a patch to interact with its copy, prevents dimerization.

Appendix C: Determination of the phase diagram

The phase diagram is determined using specialized Monte Carlo (MC) techniques. Gibbs Ensemble MC simulations directly determine the coexistence densities of the metastable gas and liquid phases [47]. We simulate a total of $N = 1000$ particles for 10^6 MC cycles, where each cycle consists on average of N particle displacements, N particle rotations, $N/10$ particle swaps, and 2 volume V changes. The critical temperature T_c and density are then estimated using the law of rectilinear diameters [48].

Because the gas-liquid line is metastable, we observe that for certain models crystallization happens so quickly that we are unable to determine the gas and liquid densities. In such cases, we estimate the critical temperature as the lowest temperature at which phase separation is not observed within 10^6 MC cycles.

To determine the fluid-solid coexistence curve, we integrate the Clausius-Clapeyron equation starting from one coexistence point using a fourth-order predictor-corrector algorithm [5]. The coexistence point is determined using free energy calculations. The free energy of the fluid is computed using thermodynamic integration from the free energy of an ideal gas [49]. The free energy of the crystal is determined using for reference the Einstein crystal with fixed center of mass [50] with Hamiltonian

$$H^{\text{Ein}}(\Xi_{\text{trans}}, \Xi_{\text{or}}) = \Xi_{\text{trans}} \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_{i,0})^2 + \Xi_{\text{or}} \sum_{i=1}^N f(\theta_i, \phi_i, \chi_i),$$

where $f(\theta_i, \phi_i, \chi_i) = 1 - \cos(\psi_{i,1}) + 1 - \cos(\psi_{i,2})$, $(\theta_i, \phi_i, \chi_i)$ are the Euler angles describing the orientation of particle i and $\psi_{i,j}$ is the angle formed between the vector defining patch j of particle i and the corresponding vector in the Einstein crystal. As explained in Ref. [49], the free energy of the reference Einstein crystal can then be written as

$$a_{\text{Ein}}^{\text{COM}} = a_{\text{trans}}^{\text{COM}} + a_{\text{or}}^{\text{COM}}, \quad (\text{C1})$$

where

$$\beta a_{\text{trans}}^{\text{COM}} = -\frac{3}{2} \frac{N-1}{N} \ln \left(\frac{\pi}{\beta \Xi_{\text{trans}}} \right) - \frac{3}{2N} \ln N \quad (\text{C2})$$

and

$$\beta a_{\text{or}}^{\text{COM}} = -\ln \left\{ \frac{1}{8\pi^2} \int d\theta \sin(\theta) d\phi d\chi \exp[-\beta \Xi_{\text{or}} f(\theta, \phi, \chi)] \right\}.$$

TABLE B.1. Geometry parameters: the triplets of numbers represent the (x,y,z) coordinates of each patch

patch ₁	patch ₂	patch ₃	patch ₄	patch ₅	patch ₆	γ
-0.8036	0.8036	-0.5186	-0.5186	0.3081	0.3981	
-0.5042	-0.5042	0.2731	0.2371	-0.8084	0.8084	0.0172
-0.3163	-0.3163	0.8102	-0.8102	0.5016	0.5016	
-0.7904	0.7904	-0.4227	-0.4227	0.3571	0.3571	
-0.5184	-0.5184	0.2807	0.2897	-0.776	0.776	0.0217
-0.3263	-0.3263	0.8617	-0.8617	0.5199	0.5911	
-0.7191	0.7191	-0.3813	-0.3813	0.3475	0.3475	
0.5659	0.5659	-0.1669	-0.1669	-0.7668	0.7668	0.0381
-0.4032	-0.4032	-0.9093	0.9093	0.5397	0.5397	
-0.6167	0.6167	0.6103	0.6103	-0.006	-0.006	
0.6521	0.6521	-0.3099	-0.3099	-0.9579	0.9579	0.0631
-0.441	-0.441	0.729	-0.729	0.2871	0.2871	
-0.9042	0.9042	-0.6276	-0.6276	0.2859	0.2859	
-0.3335	-0.3335	0.5386	0.5386	-0.9074	0.9074	0.0787
-0.2669	-0.2669	0.5622	-0.5662	0.3081	0.3081	

TABLE B.2. Energy parameters

ϵ_1	ϵ_2	ϵ_3	ζ
2.	2.	2.	0.00
1.2462	2.5482	2.2056	0.11
2.1	2.9066	0.9934	0.21
0.4854	2.8266	2.688	0.33
3.5756	0.2037	2.2207	0.49
3.	3.	0.	0.50
3.96	1.8	0.24	0.55
4.32	1.5	0.18	0.64
4.8	0.6	0.6	0.79

The calculation of $a_{\text{trans}}^{\text{COM}}$ is straightforward, but that of $a_{\text{or}}^{\text{COM}}$ requires either a tedious numerical integration, as in Ref. [51], or an analytical approximation. We opt for the latter using a saddle point approximation, which

is very efficient for high values of $\beta\Xi_{\text{or}}$ such as those used here, because the integrand is sharply peaked. Defining $(\theta_0, \phi_0, \chi_0)$ as the reference orientation in the Einstein crystal and changing variable $\alpha = (\cos(\theta), \phi, \chi)$ gives

$$\int d\theta \sin(\theta) d\phi d\chi \exp[-\beta\Xi_{\text{or}} f(\alpha)] = \int d\alpha \exp[-\beta\Xi_{\text{or}} f(\alpha)] \approx \frac{\exp[-\beta\Xi_{\text{or}} f(\alpha_0)] (2\pi)^{3/2}}{(\beta\Xi_{\text{or}})^{3/2} \det(H[f(\alpha_0)])^{1/2}} = \frac{(2\pi)^{3/2}}{(\beta\Xi_{\text{or}})^{3/2} \det(H[f(\alpha_0)])^{1/2}},$$

such that

$$\beta a_{\text{or}}^{\text{COM}} \approx \frac{3}{2} \ln(\beta\Xi_{\text{or}}) + \frac{1}{2} \ln\{8\pi \det(H[f(\alpha_0)])\}, \quad (\text{C3})$$

where $\det(H[f(\alpha_0)])$ is the determinant of the Hessian of function f computed at α_0 . The analytical expression of $f(\alpha)$ is reported below defining $y = \cos(\theta)$ and ζ as the angle between the vectors identifying patch 1 and 2, which does not depend on the orientation of the particle.

$$\begin{aligned} f(\alpha) = & \frac{1}{2} \left\{ 4 - 3\sqrt{1-y^2} \cos(\phi) \cos(\phi_0) \sin(\theta_0) - \sqrt{1-y^2} \cos(\chi) \cos(\chi_0) \sin(\theta_0) - 2\sqrt{1-y^2} \cos(\phi) \cos(\zeta) \sin(\phi_0) \sin(\chi_0) \sin(\zeta) \right. \\ & - 2y \cos(\chi_0) \cos(\zeta) \sin(\theta_0) \sin(\zeta) + 2y \cos(\phi) \cos(\chi) \sin(\phi_0) \sin(\chi_0) \sin^2(\zeta) - 2y \cos(\phi) \cos(\phi_0) \sin(\chi) \sin(\chi_0) \sin^2(\zeta) \\ & - \sqrt{1-y^2} \cos(\chi) \cos(\chi_0) \sin(\theta_0) \sin^2(\zeta) + y \cos(\phi) \cos(\phi_0) \cos(\chi) \sin(\theta_0) \sin(2\zeta) + \cos(\phi) \sin(\phi_0) \sin(\chi) \sin(\theta_0) \sin(2\zeta) \\ & + \cos(\theta_0) [-3y + \cos(\phi_0) \cos(\chi_0) \sin(\phi) \sin(\chi) + \cos^2(\zeta) (-y + y \cos(\chi) \cos(\chi_0) \sin(\phi) \sin(\phi_0) - \cos(\phi_0) \cos(\chi_0) \sin(\phi) \sin(\chi))] \\ & - 2\sqrt{1-y^2} \cos(\chi) \cos(\zeta) \sin(\zeta) + y \sin^2(\zeta) + \cos(\phi_0) \cos(\chi_0) \sin(\phi) \sin(\chi) \sin^2(\zeta) - 2 \cos(\phi) \cos(\chi_0) \sin(\phi_0) \sin(\chi) \sin^2(\zeta) \\ & - y \cos(\chi) \cos(\chi_0) (2 \cos(\phi) \cos(\phi_0) \sin(\zeta) + \sin(\phi) \sin(\phi_0) (1 + \sin^2(\zeta))) + \sqrt{1-y^2} \cos(\phi) \cos(\phi_0) \cos(\chi_0) \sin(2\zeta) \\ & + \sqrt{1-y^2} \cos(\chi_0) \sin(\phi) \sin(\phi_0) \sin(2\zeta) \left. \right\} + \sin(\phi) [-\sin(\phi_0) (2 \sin(\chi) \sin(\chi_0) \sin^2(\zeta) \\ & + \cos(\phi_0) (-2 \cos(\zeta) \sin(\chi) \sin(\theta_0) \sin(\zeta) + \sin(\chi_0) (-2y \cos(\chi) \sin^2(\zeta) + \sqrt{1-y^2} \sin(2\zeta)))]. \end{aligned}$$

Once the free energy of the reference crystal is known, the free energy of the actual crystal is obtained following a standard protocol [51]. Several simulations along an isobar starting from the fluid and from the crystal are then necessary to determine the temperature at which the chemical potential of the two phases coincides [49, 51].

Appendix D: Wertheim's perturbation theory

According to Wertheim's theory [35, 36], the free energy of the fluid can be approximated by the hard sphere free energy plus a *bond free energy* correction

$$a_f = a_{\text{HS}} + a_{\text{bond}}, \quad (\text{D1})$$

where

$$\beta a_{\text{bond}} = \sum_{a \in \Gamma} \left(\ln X_a - \frac{X_a}{2} \right) + \frac{m}{2}. \quad (\text{D2})$$

Here m is the total number of attractive sites and X_a is the probability that the molecule is not bonded at site a .

Similarly, the chemical potential is given by

$$\beta \mu_f = \beta a_f + \frac{\beta p}{\rho} = \beta a_{\text{HS}} + \beta a_{\text{bond}} + \frac{\beta p_{\text{HS}}}{\rho} + \frac{\beta p_{\text{bond}}}{\rho}, \quad (\text{D3})$$

where

$$\beta p_{\text{bond}} = \rho^2 \sum_{a \in \Gamma} \left(\frac{\partial X_a}{\partial \rho} \right) \left(\frac{1}{X_a} - \frac{1}{2} \right). \quad (\text{D4})$$

In the solid, $\beta a_s \approx \beta \mu_s$, because the ratio $\frac{\beta p_s}{\rho_s}$ is small [27, 52]. The energetic contribution to the free energy is the energy of the fully-bonded system $-\beta \epsilon_{\text{tot}}$, while the entropic term is approximated using the range of interaction and the width of the patches [27],

$$\beta \mu_s = \beta a_s = -3 \ln(\lambda - 1) - \ln \left(\frac{\delta^3}{\pi^2} \right) - \beta \epsilon_{\text{tot}}. \quad (\text{D5})$$

At coexistence, temperature, pressure and chemical potential of the fluid and the solid phase have to be identical. The pressure of the solid is once again ignored, so the only remaining constraint for coexistence is $\beta_c \mu_f = \beta_c \mu_s$. Using the equations above, it follows that

$$\begin{aligned} \beta_c a_{\text{HS}} + \beta_c a_{\text{bond}} + \frac{\beta_c p_{\text{HS}}}{\rho} + \frac{\beta_c p_{\text{bond}}}{\rho} &= \\ &= -3 \ln(\lambda - 1) - \ln \left(\frac{\delta^3}{\pi^2} \right) - \beta_c \epsilon_{\text{tot}}. \end{aligned} \quad (\text{D6})$$

As the hard-sphere system is temperature independent, it holds that

$$\beta_c \left(a_{\text{bond}} + \frac{p_{\text{bond}}}{\rho} + \epsilon_{\text{tot}} \right) = C(\rho), \quad (\text{D7})$$

where

$$C(\rho) = -\beta a_{\text{HS}} - \frac{\beta p_{\text{HS}}}{\rho} - 3 \ln(\lambda - 1) - \ln \left(\frac{\delta^3}{\pi^2} \right), \quad (\text{D8})$$

which is a constant that depends only on ρ . It follows that $a_{\text{bond}} + \frac{p_{\text{bond}}}{\rho} + \epsilon_{\text{tot}}$ represents a good rescaling factor for the temperature to obtain the master solubility line across the different models.

Appendix E: Grand Canonical Monte Carlo approximation for critical point

To check the position of the critical point for the extreme case of Fig. 2C ($\epsilon_1 = 4.6655$, $\epsilon_2 = 1.2908$ and $\epsilon_3 = 0.0437$), we perform Grand Canonical Monte Carlo (GCMC) simulations. Because one of the patches is markedly stronger than the others, percolation takes place at high temperature. Consequently, the simulation dynamics is slow and the estimate of the critical temperature is affected by large errors. Indeed, the GCMC simulations have a poor sampling. To obtain a better estimate of the phase diagram, we perform GCMC for systems with increasing strength of the strongest patch (keeping the other patches identical) and we fit a power-law to the resulting critical temperatures (Fig.E.1). The value of the fit for $\epsilon_1 = 4.6655$ ($T = 0.045$ in units of ϵ_{tot}) confirms the stability of the critical point with respect to the solubility line.

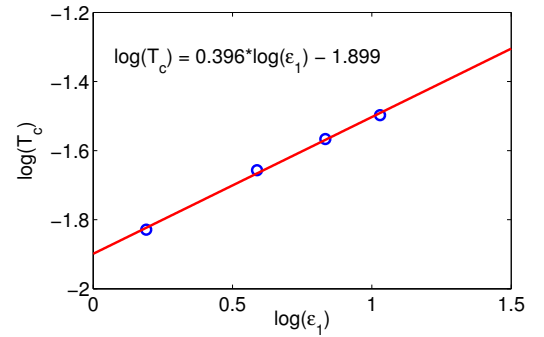


FIG. E.1. Fit of the critical point determined with GCMC simulations for systems with increasing strength of the strongest patch.

Appendix F: Determination of the percolation threshold

To determine the temperature at which percolation takes places, we run 20 *NVT* simulations with respectively $N = 2048$, $N = 4000$ and $N = 6912$ at several temperatures and $\rho = 0.2$. During the simulation, we determine the size of the biggest network defined as the

largest set of particles connected by at least two bonds. If such a cluster grows sufficiently to span the whole simulation box along one dimension within 10^5 MC sweeps, the system is deemed percolating. The probability of percolation is the fraction of simulations that show a percolating cluster. For finite-size rescaling we use the tabulated critical exponents and the standard procedure [46]. The data are in agreement with the tabulated 3D critical exponent to within 1%.

Appendix G: Definition of crystal cluster and umbrella sampling simulations

We determine the size of the crystal clusters in the simulation box following a standard procedure that defines *crystal-like bond* and *crystal-like particles* [53]. Due to the highly specific patch-patch interaction of our model, we straightforwardly define a crystal-like bond between particles 1 and 2 when they are bonded in our model: $r < \lambda\sigma$, $\theta_{1,2i} < \delta$ and $\theta_{2,2i-1} < \delta$ for some i . For $\zeta \geq 0.33$, the weakest interactions tend to be looser, so we

relax the conditions and define a bond when $r < 1.2\sigma$, $\theta_{1,2i} < 2\delta$ and $\theta_{2,2i-1} < 2\delta$. A particle is considered to be crystal-like if it has six crystal-like bonds. Two crystal-like particles belong to the same crystal cluster if a crystal-like bond connects them. Visual inspection of these “crystals” confirms that the criterion selects actual crystal clusters.

For the umbrella sampling simulations, we use a biasing harmonic potential with spring constant κ that varies between 0.07 and 0.12, depending on the model and the temperature

$$H^{\text{bias}} = \kappa(n - n_0)^2, \quad (\text{G1})$$

where n represents the size of the largest crystal cluster and n_0 is the cluster size corresponding to sampling window. The sampling windows are typically positioned every 3 particles, although sometimes denser sampling is required. The results of each simulation are then analyzed following the umbrella sampling standard protocol [53].

-
- [1] A. McPherson, *Crystallization of Biological Macromolecules* (CSHL Press, Cold Spring Harbor, 1999).
 - [2] N. E. Chayen, in *Advances in Protein Chemistry and Structural Biology*, edited by J. Andrzej (Academic Press, London, 2009), Vol. 77, pp. 1–22.
 - [3] A. George and W. W. Wilson, *Acta Crystallogr. D* **50**, 361 (1994).
 - [4] Z. S. Derewenda, *Structure* **12**, 529 (2004).
 - [5] M. H. J. Hagen and D. Frenkel, *J. Chem. Phys.* **101**, 4093 (1994).
 - [6] D. Rosenbaum, P. C. Zamora, and C. F. Zukoski, *Phys. Rev. Lett.* **76**, 150 (1996).
 - [7] P. R. ten Wolde and D. Frenkel, *Science* **277**, 1975 (1997).
 - [8] J. Janin, *Prog. Biophys. Mol. Biol.* **64**, 145 (1995).
 - [9] O. Carugo and P. Argos, *Protein Sci.* **6**, 2261 (1997).
 - [10] R. P. Bahadur, P. Chakrabarti, F. Rodier, and J. Janin, *J. Mol. Biol.* **336**, 943 (2004).
 - [11] P. J. Lu *et al.*, *Nature* **453**, 499 (2008).
 - [12] C. Haas, J. Drenth, and W. W. Wilson, *J. Phys. Chem. B* **103**, 2808 (1999).
 - [13] A. Lomakin, N. Asherie, and G. B. Benedek, *J. Chem. Phys.* **104**, 1646 (1996).
 - [14] A. Lomakin, N. Asherie, and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9465 (1999).
 - [15] R. A. Curtis, H. W. Blanch, and J. M. Prausnitz, *J. Phys. Chem. B* **105**, 2445 (2001).
 - [16] J. J. McManus *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16856 (2007).
 - [17] C. Gögelein *et al.*, *J. Chem. Phys.* **129**, 085102 (2008).
 - [18] E. Bianchi, R. Blaak, and C. N. Likos, *Phys. Chem. Chem. Phys.* **13**, 6397 (2011).
 - [19] A. McPherson, *Introduction to macromolecular crystallography, Macromolecular crystallography* (Wiley-Liss, Hoboken, N.J., 2003).
 - [20] G. Pellicane, G. Smith, and L. Sarkisov, *Phys. Rev. Lett.* **101**, 248102 (2008).
 - [21] N. Dorsaz, L. Filion, F. Smallegange, and D. Frenkel, *Farad. Disc.* **159**, 9 (2012).
 - [22] D. Fusco, J. J. Headd, A. De Simone, and P. Charbonneau, *arXiv:1206.6332* (2012).
 - [23] S. C. Glotzer and M. J. Solomon, *Nat. Mater.* **6**, 557 (2007).
 - [24] S. C. Glotzer, *Science* **306**, 419 (2004).
 - [25] F. J. Martinez-Veracoechea, B. M. Mladek, A. V. Tkachenko, and D. Frenkel, *Phys. Rev. Lett.* **107**, 045902 (2011).
 - [26] Y. Wang *et al.*, *Nature* **491**, 51 (2012).
 - [27] R. P. Sear, *J. Chem. Phys.* **111**, 4800 (1999).
 - [28] N. Kern and D. Frenkel, *J. Chem. Phys.* **118**, 9882 (2003).
 - [29] see Supplementary Material for details on the model (Appendix A), the geometry asymmetry parameter (Appendix B), phase diagram calculations and simulation procedure (Appendix C and E), Wertheim adhesion parameter derivation (Appendix D), crystal cluster and liquid cluster definition (Appendix F).
 - [30] M. G. Noro and D. Frenkel, *J. Chem. Phys.* **113**, 2941 (2000).
 - [31] The unit temperature T (and its inverse, β) is set by $\epsilon_1 + \epsilon_2 + \epsilon_3 = \epsilon_{\text{tot}}$ with Boltzmann’s constant $k_B = 1$. The geometrical orientation varies from one pair of patches to the next. Although these models can crystallize in other lattices at high temperatures and densities, these structures are not relevant for protein crystallization.
 - [32] S. W. Wukovitz and T. O. Yeates, *Nat. Struct. Mol. Biol.* **2**, 1062 (1995).
 - [33] W. L. Miller and A. Cacciuto, *J. Chem. Phys.* **133**, 234903 (2010).
 - [34] At high energy asymmetries, the trend is less clear be-

- cause the identity of the patches matter. Shuffling the patch identities changes T_c by up to 10%.
- [35] M. S. Wertheim, J. Stat. Phys. **35**, 19 (1984).
 - [36] M. S. Wertheim, J. Stat. Phys. **35**, 35 (1984).
 - [37] G. Foffi and F. Sciortino, J. Phys. Chem. B **111**, 9702 (2007).
 - [38] We assume that the simulated models critical density $\rho_c \approx 0.2$ is constant, which is numerically reasonable. Wertheim theory, however, notoriously underestimates the critical density for patchy particle models [54].
 - [39] P. G. Vekilov and A. A. Chernov, Solid State Physics **57**, 1 (2002).
 - [40] N. Asherie, Methods **34**, 266 (2004).
 - [41] F. Bonnet, D. Vivars, C. Robert, and N. Colloch, J. Cryst. Growth **232**, 330 (2001).
 - [42] J. Blouwolff and S. Fraden, J. Cryst. Growth **303**, 546 (2007).
 - [43] T. K. Haxton and S. Whitlam, Soft Matter **8**, 3558 (2012).
 - [44] L. Redecke *et al.*, Science (2012).
 - [45] J. M. Tavares, P. I. C. Teixeira, and M. M. Telo da Gama, Phys. Rev. E **81**, 010501 (2010).
 - [46] D. Stauffer and A. Aharony, *Introduction to percolation theory* (Taylor & Francis, London, 1992).
 - [47] A. Z. Panagiotopoulos, Mol. Phys. **61**, 813 (1987).
 - [48] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, London, 2001).
 - [49] C. Vega, E. Sanz, J. L. F. Abascal, and E. G. Noya, J. Phys.-Condens. Mat. **20**, 153101 (2008).
 - [50] D. Frenkel and A. J. C. Ladd, J. Chem. Phys. **81**, 3188 (1984).
 - [51] F. Romano, E. Sanz, and F. Sciortino, J. Chem. Phys. **132**, 184501 (2010).
 - [52] G. Jackson, W. G. Chapman, and K. E. Gubbins, Mol. Phys. **65**, 1 (1988).
 - [53] I. Saika-Voivod, F. Romano, and F. Sciortino, J. Chem. Phys. **135**, 124506 (2011).
 - [54] E. Bianchi *et al.*, Phys. Rev. Lett. **97**, 168301 (2006).